DOCUMENT RESUME

ED 258 191 CS 208 936

AUTHOR	Jackson, Rex
TITLE	The Effects of the Mason Thinking and Writing Program on the Skills of Second Grade Children. Technical Report.
INSTITUTION	Applied Educational Research, Inc., Princeton, NJ.
REPORT NO	AER-85-1
PUB DATE	Feb 85
NOTE	81p.; Funding provided by the Mason Early Education Foundation.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Elementary Education; *Expository Writing; *Program Effectiveness; *Program Evaluation; Sentences; Writing Exercises; *Writing Improvement; Writing Processes; Writing Research; Writing Skills
IDENTIFIERS	*Thinking Skills; *Writing Programs

ABSTRACT

A study assessed the effects of the Mason program (a supplementary language arts program in which expository writing is introduced) on writing skills, such as writing mechanics, expression, and related thinking skills. These skills were measured by standardized tests and by actual writing performance. Subjects, approximately 400 second grade students divided into a treatment group (in the Mason program) and a control group, were tested at the beginning and end of instruction. Program students were asked to write in such a way that a humorous imaginary character named Mr. Magargle (who is easily confused and who is always in need of help) can understand written explanations or instructions. The lessons included specific instruction and practice in exercising a variety of thinking skills important in expository writing, including classification skills, presenting a logical sequence of ideas, and distinguishing between the necessary and the unnecessary. Students were taught to reread and correct their own work. Results indicated that for most writing measures, statistically significant group differences favoring the treatment group were found, especially in sentence quality and in the overall quality of responses to writing exercises. Also, the great majority of experienced teachers thought that the Mason program had a moderately or strongly positive effect on student writing. (EL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

AER Report 85-1
February, 1985

ED258191

THE EFFECTS OF THE MASON THINKING AND WRITING PROGRAM ON THE SKILLS OF SECOND GRADE CHILDREN

Technical Report

Rex Jackson

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Rex Jackson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS208 936

Applied Educational Research, Inc.

THE EFFECTS OF THE MASON THINKING AND WRITING PROGRAM ON THE SKILLS OF SECOND GRADE CHILDREN

Technical Report

Rex Jackson

Applied Educational Research, Inc.
Princeton, New Jersey
February, 1985

ACKNOWLEDGEMENTS

Many individuals contributed to the research project described in this report.

Financial support for this research was provided by the Mason Early Education Foundation. In addition, Mary Mason and her staff provided assistance with many aspects of this work. Appreciation is due to Mary Mason, in particular, for her continuing advice and support, and for her insistence that this project constitute an independent appraisal of the effects of the Thinking and Writing program. The author of this report had complete freedom to design and conduct this research and to interpret the results in accordance with his own professional judgments.

Special thanks are due to the twenty-two second-grade teachers whose classes took part in this study (and to the students) -- for accommodating to instructional changes required by the study design and for administering (and taking) special tests and exercises.

Thanks are due also to the administration of the Putnam City, Oklahoma school district -- to Ralph Downs, Superintendent of Schools, for his support of this research, and to Harold Greenwood, Assistant Superintendent for Curriculum, for his support and for his continuing assistance and counsel during the course of the study.

A very large debt of gratitude is owed to Lenora Thorsen, Coordinator of Language Arts for the Putnam City District, who was responsible for carrying out or coordinating many essential tasks including initial identification of eligible classes, making all necessary local arrangements for getting the project underway, monitoring progress

during the course of the study, and coordinating post-testing. Without her help in managing virtually all aspects of the study in the schools and in the district office, this project could not have been carried out.

The contributions of a number of individuals who contributed to other aspects of this work are also acknowledged with gratitude: Kalle Gerritz who provided assistance in development of special thinking and writing exercises; Susan Jackson who provided counsel throughout the study and who coordinated all scoring and data preparation; Geraldine May who directed the reading of student papers; Jermaine Andrews, Anita Loerke, Patricia Mills, Nancy Osborn, Martha Palmieri, and Susan Thomas who served as readers; Mait Jones and Alka Naithani who scored the multiple choice post-tests and who coded student papers; Kelly Martin and Jacquelin Masotti who performed the content analysis and who developed the other special counts and scores for student papers.

Rex Jackson
February 1985

CONTENTS

Summary	1	
The Instructional Program	7	
The Study Plan	9	
Study Variables	17	
1. Pre-tests		17
2. Post-tests		19
Findings	25	
1. Results of Analyses of Student Data -- Writing Scores		25
2. Results of Analyses of Student Data -- Content Indicators		35
3. Responses to the Teacher Questionnaire		38
References	43	
Appendix	45	
Tables	51	
Figures	73	

SUMMARY

This report presents a number of conclusions concerning the impact on children's skills of an instructional program in expository writing for young children -- the Thinking and Writing program developed and published by the Mason Early Education Foundation. These conclusions are based on the results of an empirical curriculum evaluation study of this program conducted by Applied Educational Research, Inc. -- a study involving both pre-testing and measurement of a variety of learning outcomes for groups of second-grade children in twenty-two classrooms in seven schools in the Putnam City, Oklahoma school district -- an independent district within the boundaries of Oklahoma City. Half of the twenty-two classes were assigned at random to a "treatment" group in which the Mason program was taught during the fall of 1983, replacing a a portion of the language arts curriculum regularly used in this district. The remaining classes constituted a control group in which the regular language arts program was taught during this semester.

All children were administered a standardized achievement test battery -- the Comprehensive Tests of Basic Skills (CTBS) -- at the beginning of the school year. In December 1983, when treatment classes had completed the program, several sections of a parallel form of the CTBS and five specially developed writing exercises were administered in all participating classes. In the analyses, differences between children in the treatment and control groups in their performance on these outcome measures are examined after taking into account scores on the language arts sections (reading, spelling, and language) of the pre-test battery. This statistical control for pre-existing differences on a standardized test measuring relevant skills, along with the random

assignment of classes to groups, provides a reasonable basis for concluding that significant group differences on post-tests can be attributed to the effects of the Mason program and not to other factors.

Outcome measures -- scores on the CTBS post-tests and scores or ratings based on responses to the special exercises -- included assessments of the following skills or attributes of writing:

- Recognition of correct capitalization and punctuation, grammar and sentence structure
- Overall judged quality of responses to assigned writing tasks
- Use of "uncommon" vocabulary in writing (words not on a list of 200 common words)
- Use of correct sentences in writing
- Spelling

In addition, content analyses of responses to the writing exercises were performed yielding indicators of (a) correctness in completing the classification or sequencing tasks implicit in several of the writing exercises, and (b) use of specific vs. general attributes in classification.

Certain measures of possible "side effects" (outcomes not reflecting program objectives) were also obtained. For example, the amount of writing produced in response to free response exercises was measured by counting numbers of words and sentences used, even though the program does not aim to encourage children to respond at length to expository writing assignments.

For most writing outcome measures examined, statistically significant group differences favoring the treatment group were found. Furthermore, in nearly every case in which group differences were found, significant group by pre-test interactions were found, indicating that treatment effects are not constant for all students. Average treatment/control group differences were consistently found to be larger for students with low initial scores on the CTBS pre-test than for students with higher pre-test scores.

Figure 1 and Figure 2 show the typical pattern of results for the two post-test measures showing the largest differences: (1) total holistic scores (the total of overall quality ratings for responses to three writing exercises), and (2) percent of sentences written in response to the special exercises that were classified as correct sentences (complete sentence, initial capital, end punctuation). As the figures indicate, for most pre-test values, expected outcomes on these measures are higher for the treatment than for the control group, and the expected differences are especially large for low scoring students. On the average, students in classes completing the Mason program -- particularly students with low test scores at the beginning of the year -- write responses to brief writing exercises that are judged by qualified readers to exhibit better writing quality and that exhibit more correct sentences than papers written by comparable students in classes not using the program.

Patterns like those depicted in Figures 1 and 2 were found for a number of other measures of writing skills, although differences on other variables were generally not as large. The consistency with which

this pattern was found in our results strongly suggests that students who begin second grade with relatively weak skills (as measured by reading, spelling, and language sections of a standardized test) profit to a particularly large degree from the kind of instruction that makes up the Mason Thinking and Writing program. The estimated effects for low scoring students on a number of outcome measures were clearly large enough to be of practical importance, as well as statistically significant.

Because of the relatively large effects for initially low-scoring students, we found for most post-test measures less variation among treatment group students than among control group students. It is often found in curriculum evaluation studies that effective programs increase the gap between the most and least able students. Contrary to that common result, our findings suggest the Mason program is particularly helpful in improving writing skills of less well prepared students, and that it may help to close the gap between students who begin second grade near the top and near the bottom of their classes.

As indicated above, the largest positive effects were found for a measure of sentence quality and for the overall judged quality of responses to writing exercises. We also found differences favoring the treatment group (or lower scoring students in that group) in the proportion of uncommon words used in writing and in spelling (measured both by the CTBS Spelling test administered as a post-test and by the percentage of words in writing exercise responses that were correctly spelled). These findings supported expectations that the regular practice in writing experienced by students in the Mason program would result in improved spelling and in use of somewhat richer vocabulary in writing. In general, students in the two groups used about the same number of

words and sentences in responding to the writing exercises, though for students at about the 20th percentile and below on the pretest, statistically significant, but relatively small treatment effects were found, with children in the treatment group producing somewhat longer responses.

Two post-tests measuring aspects of writing skills showed no significant differences -- the Language Mechanics and Language Expression sections of the CTBS -- tests which were made up of multiple-choice questions requiring recognition of correct capitalization and punctuation or grammar and sentence structure. Since these are skills emphasized in the regular language arts program in use in the control classes, this was not a surprising result. In general, the effects of the Mason program were evident in production tasks -- in measures of actual writing -- rather than in recognition tasks dealing primarily with mechanics of writing.

Analyses of several indicators developed through content analyses of student responses also found statistically signficant positive treatment effects. Students who had completed the Mason lessons showed higher levels of success in stating classification rules and in providing a logical sequence of steps in response to exercises calling for these thinking skills. Treatment group students also tended to use specific attributes of objects more often in stating classification rules, and they used general descriptions or superordinate categories less frequently -- results that were anticipated because of the focus on specificity in completing such tasks in a number of the Mason thinking and writing lessons.

Finally, at the conclusion of the study, teachers in all participating classes who had taught the program in this or in prior years were asked for their judgments of the effects of the Mason program on children's skills and attitudes. Paralleling our findings based on analyses of the student data, teachers' perceptions of the program were highly positive. For a large number of skills and attitudes that the program was thought to influence, the great majority of experienced teachers indicated that they thought the Mason program has a moderate or strong positive effect.

THE INSTRUCTIONAL PROGRAM

The Mason Thinking and Writing program (Mason Early Education Foundation, 1981) is an introduction to expository writing. It is intended to supplement the instruction in language arts that children in elementary school generally receive. The second-grade program consists of thirty-five lessons, one lesson to be taught each week. Each lesson is divided into two half-hour parts which are to be taught on two successive days. In a typical lesson, the first part encourages thinking about a topic or problem and may involve class discussion as well as worksheets for the children to complete. In the second part of a typical lesson, the children are asked to write and to correct one or two sentences, or a brief paragraph, on the topic or problem.

Children are asked to write in such a way that a humorous imaginary character named Mr. Magargle (who is easily confused and who is always in need of help) can understand written explanations or instructions. The lessons include specific instruction and practice in exercising a variety of thinking skills important in expository writing, including classification skills, presenting a logical sequence of ideas, and distinguishing between what is necessary and what is unnecessary.

The children are taught to reread and correct their own work by putting check marks in a series of color-coded circles provided on each paper. A red circle reminds the child to check for sentence completeness, capitalization, and punctuation. A green circle is a reminder to stick to the assigned point. A blue circle is a reminder to check that the sequence of instructions or events presented in a paper is correct.

The program provides teachers with a set of well-defined and structured lessons that are designed to introduce young children to certain fundamentals of good expository writing, to develop related thinking skills, and to provide a foundation for future development. A more detailed outline of the aims and features of the second grade level of the Thinking and Writing program -- a description provided by the program developer -- is given as an appendix to this report.

THE STUDY PLAN

The evaluation study described here was carried out in the Putnam City, Oklahoma school district in 1983-84. The Mason Thinking and Writing program had been used throughout the Putnam City district at the second-grade level in the 1981-82 and the 1982-83 school years. Teachers in the district were, therefore, familiar with the program. Furthermore, the program appeared to have strong support among many teachers and among district administrators. Thus, this district appeared to offer several advantages for a study of this kind. By conducting the research in this setting, we would be less subject to the distorting effects that arise when an entirely new and unfamiliar program is evaluated in its first year of use, and we could expect to receive needed support and cooperation from teachers and administrators.

The principal purpose of the study was to assess the effects of the Mason program on student skills -- mechanics of writing, expression, and related thinking skills as measured by both standardized tests and by actual writing performance. Three features were judged to be essential for a study to yield dependable inferences about how effective the Mason program is in helping to develop these skills:

1. Outcome measures used in the study had to include direct and valid assessments of the particular skills that the Mason program is intended to strengthen. In particular, since a major aim of the program is to give children practice in the necessary skills to write a brief expository paragraph in response to a given topic, it was judged important to include actual writing exercises as outcome measures for the project,

and not to rely solely on standardized, multiple-choice tests, although such tests were given in addition to writing exercises.

2. The study had to yield appropriately controlled comparisons of students who had received instruction in the Mason program and those who had not, so that differences in outcomes for these groups could be interpreted as effects of the instructional program and not attributed to other factors. Thus, it was necessary to have a treatment group of students taught the Mason lessons, and a control group of otherwise comparable students taught only the regular language arts program used in the district (but not the Mason lessons).

3. The Mason program had to be fully implemented and taught to its conclusion in the classes in the treatment condition, and not used at all in classes in the control condition during the course of the study.

These considerations argued for conducting the study within a manageable subset of schools and classrooms within the total district, so that the participating classes could be regularly monitored to insure that the program was being fully implemented in the treatment classes and so that testing and other special research requirements could be appropriately coordinated and managed.

A total of 26 classes in 7 schools were initially identified in the late summer of 1983 as candidates for participation in the study. The 26 classes were divided into 13 pairs, so that each pair included two classes from the same school. Then, one member of each pair was

assigned at random to the treatment condition and the other to the control condition.

Since the average number of students per class was expected to be between 20 and 25, we expected more than 500 students to be enrolled in these selected classes. Allowing for attrition both at the class and student level, our aim was to have a final sample of approximately 400 students with about half assigned to each condition -- numbers judged to be large enough to yield relatively stable estimates of effects while still being manageable.

If any class (treatment or control) had to be dropped from the study, the class with which it had been originally paired and which had been assigned to the alternate condition was also dropped. In fact, two pairs (four classes) were lost. In one case, changes early in the school year in assignment of children to classes in one school and reassignment of one teacher who was to have participated to a third grade class made it necessary to drop a pair. In the second case, a pair was dropped because one class assigned to the treatment condition had not nearly completed the thirty-five Mason lessons by the time post-testing was to begin. Thus, the final sample included 22 classes (11 treatment and 11 control). Within these 22 classes, a total of 383 students had complete data on the major study variables -- 197 students in treatment classes and 186 in control classes.

Normally, the Mason program is taught during the course of a full year in about two half-hour periods per week as a supplement to the regular language arts curriculum. Because district administrators believed strongly in the value of the program and were understandably

unwilling to deny it to any children for a full year for research purposes, special arrangements were needed to permit a controlled study while giving all children access to instruction in the program under study. Accordingly, it was agreed that all 35 lessons that constitute the Mason Thinking and Writing program at the second-grade level would be taught in the treatment classes in the fall of the year, ending in early December 1983. In the control classes, instruction in the Mason program was delayed until the second half of the year, beginning in January 1984. Thus, although all students in participating classes ultimately received this instruction (either in the first half or the second half of the year), we were able to obtain a controlled comparison of students taught and not taught the program by conducting post-testing in mid-December 1983.

As a consequence of these arrangements, the results of this study are estimates of the effects of the instructional program taught in a different fashion from the way the program was designed to be used. It is possible that different results would have been obtained had we been able to conduct a comparable study over a longer period, with the Mason lessons taught (as is normally the case) about twice a week for a full year. While we have no way of judging from our student data whether the results would be different, when we surveyed teachers in participating classes who had taught the Mason program both in this year and in previous years, they favored by about two-to-one teaching the program throughout the year, and they indicated that they believed the program was more effective when used in a less concentrated fashion over a longer period.

Because teaching of the Thinking and Writing program was concentrated in the first half of the year for the treatment classes, a major portion of the language arts curriculum that is normally taught throughout the year in the Putnam City district was deferred until the second half of the year in these classes. The specific curriculum units which were used in the control classes, but not in the treatment classes, during the first half of the year were the first three units and a part of the fourth unit of the Language for Daily Use program (Strickland, 1983). The focus of these units is on several aspects of language mechanics including, for example:

-- parts of sentences; beginning capitalization and end punctuation; combining sentences

-- nouns; names, titles, and dates; noun forms -- plurals and compound words; capitalization and punctuation in relation to names and dates

-- verbs, present and past tense, forms of be and have, contractions

Thus, the specific contrasts presented by this study are for groups of children taught during a half-year period (a) the Mason Thinking and Writing program (in the treatment classes), and (b) a set of commonly used instructional units covering aspects of language mechanics (in the control classes). Apart from this difference, the two groups followed a common curriculum using the same materials in spelling, reading, and in other subjects.

As indicated above, participating classes were assigned at random to the two conditions after matching by school. As is normally the case in studies of this kind, it was not feasible to randomly assign individual students to the two conditions. Since the sampling units consisted of only 22 intact classes, there was a significant possibility that our estimates might be biased due to chance assignment of a preponderance of classes with initially brighter or better prepared students to one of the conditions. To account for this possibility, we obtained pre-test scores for individual students on the Comprehensive Tests of Basic Skills (CTBS) published by the California Test Bureau, which were given as part of the regular school testing program at the beginning of the school year.

In fact, average pre-test scores of control group students on relevant sections of the CTBS were found to be below those of the treatment group by a relatively small, but statistically significant amount. We accounted for these differences in our analyses, and all conclusions reached here are based on analyses of differences between treatment and control groups on outcome measures after controlling statistically for initial pre-test differences on several sections of the CTBS. It should be noted that since the actual factors that may have been at work in assignment of students to classes are not measured by our pretest, we cannot in theory rely entirely on analyses that use the CTBS pretest to control for pre-existing differences to yield unbiased results. As a practical matter, however, the absence of a purposeful assignment of students to classes, random assignment of classes to treatments after matching by school, and control for initial

individual differences on the CTBS give us considerable confidence that post-test differences can be attributed to effects of the treatment.

The basic schedule and plan of the study can be summarized as follows:

1. Late August 1983

   -- Assign classes to treatment and control conditions.

   -- Administer full CTBS battery as pre-test to all students.

2. September to December 1983

   -- For 11 classes in treatment group: Teach all 35 Mason Thinking and Writing lessons (about four-to-five half-hour periods per week) in place of a portion of the regular language arts program.

   -- For 11 classes in control group: Teach only the regular language arts program during this period.

3. Mid-December 1983

   -- Administer post-tests: alternate forms of three CTBS sections and special thinking and writing exercises.

As indicated above, this pre-test/post-test design with random assignment of classes to groups is one that permits reasonable interpretation of post-test differences as treatment effects. Another major strength of this study is its use of post-tests that were designed to measure the particular skills that the instructional program is designed to improve, rather than relying solely on existing measures.

The study also has certain limitations that need to be recognized, as well. One of these -- the teaching of the program in a more concentrated form than usual -- was discussed earlier. It should also be noted that the results of the study will provide information only about short-term effects of the program on the skills measured by post-tests given immediately after the completion of the total set of lessons. The second-grade program aims, however, to foster certain skills, habits, and attitudes that will constitute a sound foundation for further development of expository writing skills (not to turn second graders into accomplished writers). Certainly, the longer term effects of a program such as this one in influencing children's abilities to handle work at the higher levels is an important consideration, though such longer-term effects are not assessed directly in this study.

At the conclusion of the study in December, we also asked teachers in participating classes to complete a brief questionnaire concerning the study and their perceptions of the Mason program. Responses to certain topics covered by this questionnaire are also summarized in this report.

STUDY VARIABLES

1. Pre-tests.

The CTBS battery administered to students in the early fall had the following structure:

Word Attack Skills
Vocabulary
Reading Comprehension

Reading--Total

Spelling

Language Mechanics
Language Expression

Language--Total

Mathematics Computation
Mathematics Concepts & Applications

Mathematics--Total

In order to provide a common pre-test measure for use across a number of analyses, scaled scores on the three major sections of the CTBS most relevant to instruction in language arts (Reading--Total, Spelling, and Language--Total) were used to control for pre-existing differences. For each student, a single pre-test score was developed by averaging his or her scores on these three sections of the test.

Separate CTBS subtest scores could have been entered in the analyses as multiple covariates. However, the various CTBS scores were found to be substantially intercorrelated among themselves. The lowest correlation found between any two CTBS subtests in Reading, Spelling and Language was 0.52. A factor analysis of all the CTBS subtest scores (omitting the total scores) indicated that 81% of the variance could be

accounted for by two factors, and that each factor beyond the first two would account for only a very small percentage of variance. The two factors extracted could be clearly interpreted as a language arts factor (with substantial and approximately equal loadings for the Reading, Spelling, and Language subtests) and a mathematics factor (with substantial loadings for the two mathematics subtests). Communalities for the language arts subtests ranged from 0.75 to 0.87, values not far below the test reliabilities, indicating that these subtests have relatively small unique, non-error variance components.

Correlations betwen CTBS total scores in Reading, Spelling, and Language ranged from 0.74 to 0.80. In addition, each was expected to have at least a moderate positive correlation with each outcome measure. In a case like this, theory and experience both suggest that a simple, logically weighted linear function of scores on the relevant sections of the test (such as our choice of the mean of the three total scores on the Reading, Spelling, and Language sections of the CTBS) would have about as high a relationship with post-test scores as composites formed using partial regression weights determined in analyses of separate CTBS pre-test scores in relation to each post-test. Thus, we judged that we would lose very little by combining pre-test scores into a single measure, while gaining several important advantages. Since we intended to test for group by pre-test interactions in models accounting for a number of dependent variables, use of multiple covariates would have greatly increased the number of coefficients to be estimated. Use of a single pre-test score would alleviate possible problems in estimation due to multicollinearity and exposure to Type I error resulting from

multiple statistical tests. It would also greatly simplify the interpretation and presentation of results.

2. Post-tests.

Three sections of a parallel form of the same level of the CTBS that had been used in pre-testing were also given as post-tests in mid-December: Spelling, Language Mechanics, and Language Expression. In addition, five specially developed thinking and writing exercises were administered at this time. Three of these exercises asked students to respond briefly in writing to the following topics:

1. "First write two ways that cats and dogs are alike. Then write two ways that cats and dogs are different."

2. "Why do you think we write things? Give one or more reasons why we write."

3. "Tell in two or more sentences how you wash your face."

A variety of scores and ratings relating to different aspects of writing and thinking were developed based on student responses to these three exercises.

Each of the the three papers was read holistically (i.e, for overall quality) by qualified readers. A total of six readers were employed and each paper was read twice, with papers assigned to readers at random. Writing scores for each paper were developed (summing over the two readers), as was a total holistic score (summing over the three papers for each student).

Procedures developed and used over many years at a variety of levels by Educuational Testing Service for holistic scoring of writing samples were used for this reading, including: (a) pre-selection of a number of model papers ("range finders") to exemplify four levels of writing to be scored 1, 2, 3, and 4, (b) training of readers to the common standard set by these papers, and (c) consistency checks across readers during the reading process (using papers jointly read) to insure that common standards continued to be used. A description of the holistic reading process can be found in Fowles (1978).

In the scoring process, papers of treatment and control were identified only by individual code numbers and were mixed randomly, so that readers had no indication of the class any student belonged to or the group (treatment or control) to which the student's class belonged.

To our knowledge, this is the first time this kind of reading had been attempted for writing samples as brief as ours (averaging just over 20 words per paper) and which had been written by so young a group of writers. Analyses of the results, however, indicate that scorers were able to reliably discriminate among the papers. Inter-reader correlations were 0.75, 0.65, and 0.68 for scores assigned to the three papers (in the order listed above). The correlations between exercise total scores (summed over two readers for each paper) ranged from 0.47 to 0.51. A lower-bound estimate of the reliability of the total score is 0.74. This is the value of coefficient alpha for the total score treating each of the exercise two-reader totals as component subscores.

In addition to the holistic scores, a number of scores or counts describing other features of the papers were developed, including:

-- Number of words and sentences used

-- Number of uncommon words used -- words not on a published list of 200 common words (Sundbye, Dyck, and Wyatt, 1979)

-- Number and percent of words spelled correctly

-- Number and percent of common words spelled correctly

-- Number and percent of uncommon words spelled correctly

-- Number and percent of sentences classified as correct sentences (complete sentence, starting with a capital letter, ending with appropriate punctuation)

Content analyses of responses to several of these exercises also produced several ratings or indicators judged to be relevant to certain aims of the Mason program. Among the more important such ratings (all of which were of the yes/no or 1/0) type were: (a) Were two clearly stated similarities given in the cats and dogs exercise? (b) Were two clearly stated differences given? (c) Were specific attributes or general (superordinate) categories used as the basis of given similarities or differences? (d) Were three or more steps given in correct sequence in response to the exercise on "how you wash your face"?

Instruction in the Mason program focuses in part on encouraging thinking before writing, use of explicit rules in classification and categorization, sticking to the point, and attention to sequence. The indicators resulting from the content analyses described above are not intended to constitute comprehensive measures of the kinds of writing-related thinking skills the program intends to develop or to improve.

Instead, we considered them to be indicators only -- measures of the presence or absense in samples of writing of certain features we would expect to find more frequently in papers written by the treatment group, if the program is effective in imparting certain skills or habits of thought of the kind suggested above.

A fourth exercise, beyond the three principal writing tasks listed earlier, asked children to examine pictures of a set of 15 toys, to form two groups of three toys each that were alike in some way, and to state the feature that the toys in each group had in common. This exercise was not designed to elicit a substantial sample of writing -- only a brief sentence or phrase for each group -- and it was not read for writing quality as were the three papers described above. Responses to each part of this task were rated in the following way: 1 if a common feature of the selected group of three toys was clearly stated, otherwise 0. Thus, these ratings may also be thought of primarily as "thinking" scores, reflecting success on the classification task, though a clear statement of the classification rule was also required.

A fifth exercise asked children to read a very brief paragraph and to write a summary sentence. It was apparent on reading the papers that a large proportion of children had given responses with essentially the same content expressed in relatively similar ways. Since the samples of writing produced to this exercise were also quite brief (one sentence only), it was decided not to score responses to this exercise or to use them in the evaluation.

Given the large number of specific scores and ratings developed based on the thinking and writing exercises, it was judged important to

focus attention on those viewed as most comprehensive, important, or relevant to the aims or anticipated effects of the Mason program. In addition, we examined effects on the standardized, multiple-choice post-tests, even though they were not judged to be particularly well matched to the objectives of the instructional programs. Thus, attention centered on the following principal dependent variables:

Standardized Tests

1. CTBS Spelling
2. CTBS Language Mechanics
3. CTBS Language Expression

Overall Judgments of Writing Quality

4. Total Holistic Score (2 readers) -- Cats and Dogs Exercise
5. Total Holistic Score (2 readers) -- Why We Write Exercise
6. Total Holistic Score (2 readers) -- Wash Your Face Exercise
7. Total of Holistic Scores -- All Three Exercises

Counts and Ratings of Exercise Responses Relating to Fluency, Vocabulary, Spelling and Sentence Quality

8. Number of Words (total for 3 exercises)
9. Percent Uncommon Words (not on list of 200)
10. Percent of All Words Spelled Correctly
11. Percent of Common Words Spelled Correctly
12. Percent of Uncommon Words Spelled Correctly
13. Number of Sentences (total for 3 exercises)
14. Percent Correct Sentences (Complete, Initial Capital, End Punctuation)

Ratings Based on Content Analyses

15. Two Similarities and Two Differences -- Cats and Dogs
16. Superordinate Similarity Used -- Cats and Dogs
17. General Difference Used -- Cats and Dogs
18. Specific Difference Used -- Cats and Dogs
19. Three or More Steps in Order -- Wash Face
20. Toygroups -- Common Feature Stated for Each Group

Descriptive statistics for pre-test scores and for the 20 dependent variables listed above for a total sample of 383 students with complete data are given in Table 1.

FINDINGS

This section is devoted primarily to discussion of the results of analyses of the data collected for individual students. Following this discussion, however, certain information concerning teacher perceptions of the Mason program, based on their responses to a questionnaire, is also provided.

1. Results of Analyses of Student Data -- Writing Scores.

Analysis of covariance methods were used to estimate treatment effects after accounting for differences between the treatment and control groups measured at the outset of the study. Thus, attention is centered not on simple differences in group means, but on differences between the groups in the regressions of outcome measures on the common pre-test score (the average of initial CTBS Reading, Spelling, and Language scores). For continuous dependent variables, ordinary regression methods are used, while for the qualitative dependent variables (such as the content analysis indicators, which took values of 1 or 0), logistic regression methods are used.

The results present somewhat complex problems in interpretation, in part because of the number of variables and in part because, in many cases, outcomes were found to depend not only on pretest scores and group membership, but also on the interaction of the two. For many dependent variables, models assuming that treatment effects are constant (independent of individual's initial standing on the pretest) could not adequately account for the observed data. In such cases, it was necessary to estimate not just a single treatment effect, but different

effects for students with different pre-test scores. However, despite these complexities and the number of independent analyses carried out, the total set of findings presents a remarkably coherent picture -- a consistent pattern of statistically significant findings, nearly all of which point in a single direction.

For two of the three CTBS sections used as post-tests -- CTBS Language Mechanics and CTBS Language Expression -- no statistically significant treatment or pre-test/treatment interaction effects were found. (Apart from considerations of statistical significance, the observed group differences were also very small in practical terms.) Considering the lack of a close fit between the content of these two sections of the CTBS and the instructional program under study, the finding of negligible group differences on these two CTBS tests did not come as a surprise. The CTBS Language Mechanics test is made up of problems requiring the student to recognize errors in capitalization and punctuation. The CTBS Language Expression test consists of problems that require the student (a) to chose a completion which produces a grammatically correct sentence, or (b) to recognize which of several alternatives is a grammatically correct sentence. These tasks are similar to exercises in the instructional units stressing aspects of language mechanics that were used in the control classes but not in the treatment classes during the course of this study.

Since the Mason Thinking and Writing lessons were used in place of these curriculum units during the course of this study (rather than in addition to them as is the normal practice), students in the treatment classes undoubtedly had less practice with material resembling the content of the CTBS Language Mechanics and Language Expression tests.

However, our finding of negligible (and statistically non-significant) group differences suggests that the reduced time spent by children in treatment classes in practice on the kinds of recognition tasks included in the CTBS Language Mechanics and Language Expression tests did not in fact result in lower scores for children in the treatment classes on those multiple-choice tests.

Apart from these two CTBS test scores, twelve other post-test measures measured on at least quasi-continuous scales were analyzed by ordinary regression methods. These included the CTBS Spelling test (variable 1 on the list above), the three holistic scores for individual exercises and the total holistic score (variables 4 through 7), and a number of counts and ratings based on responses to the three major exercises (variables 8 through 14). For every one of these twelve outcome measures, the slope of the regression of outcomes on the pre-test scores was lower for the treatment than for the control group. In ten of twelve cases, the slope differences were significant at the 0.05 level. The interaction test was not significant at this conventional level for two dependent variables -- percent of all words spelled correctly and percent of uncommon words spelled correctly (t values of -1.11 and -0.87). However, the fact that the slope differences were still moderate in size and in the same direction as those for ten other post-test scores (which did exhibit significant interaction effects) persuades us to accept models including an interaction term for all of these outcome measures. Substantively, the results of all these analyses strongly indicate that the Mason Thinking and Writing program has particularly large positive effects (in comparison to the regular language arts program) on the skills of students with low initial scores.

Measures of Overall Writing Quality

The holistic scores for each of the three exercises (summed over two readers) and the total of the three exercise scores show almost the same pattern. Figure 1, which shows the treatment and control regressions of total holistic scores on the pre-test, displays this characteristic pattern -- expected outcomes that appear to be higher for the treatment group than for the control group throughout most of the range of observed pre-test scores, but marked differences in regression slopes. It appears that there are differences favoring the treatment group and particularly large positive effects for initially low scoring students.

Certain of the expected values for total holistic scores that are depicted in Figure 1 are also given in Table 2. To illustrate the magnitudes of the estimated treatment effects at different levels of the pre-test, we have chosen in Table 2 (and in a number of similar tables of results provided for other dependent variables) to display expected values of the dependent variable in the control and treatment groups at five levels of the pre-test -- scores of 438, 495, 551, 590, and 627 -- chosen because they were the 10th, 25th, 50th, 75th, and 90th percentiles in the distribution of pre-test scores for the combined (treatment and control) groups.

Figure 1 and Table 2 indicate that much larger treatment/control differences tend to be observed for students with low initial pretest scores than for students with high initial scores -- an interaction that was statistically significant for this outcome measure (t=-3.27) and

similar to interactions found, as noted above, for a variety of other measures. We have used a technique developed by Johnson and Neyman (1936), one also described by Huitema (1980, pp. 270-297), to determine at what pre-test levels the treatment/control differences can be considered significant at at least the 0.05 level. Thus, in Table 2 we also give the maximum pre-test score (and the corresponding total group percentile) for which differences can be considered statistically significant, and we have marked the illustrative differences given in the table with one asterisk (*) if they are significant at the 0.05 level and with two asterisks (**) if they are significant at the 0.01 level (as determined by the Johnson-Neyman analyses).

The column at the far right of Table 2 (and of similar tables presented later in this report) shows the ratio of the differences in expected outcome values for the treatment and control groups to the standard deviation of outcome scores for the control group -- i.e., the variation expected in the no-treatment condition. (We did not use a pooled estimate of this standard deviation, because one of the consequences of the large interaction effects is a finding of smaller standard deviations for treatment groups on the outcome measures.) These ratios, which are shown only for statistically significant differences, give some sense of the size of the differences in practical terms. As a rule of thumb, in educational evaluation studies estimated net effects of half a standard deviation or more are generally considered large enough to be clearly of practical importance.

To give a sense of the magnitude of an effect amounting to half a standard deviation, consider the accompanying change in a student's rank

on a variable that is normally distributed in some reference group of interest. Assume (as in our results) that effects of this kind are found for students who tend to have a relatively low standing on measures of academic ability. Raising performance from one standard deviation below the mean (about the 16th percentile) to one-half a standard deviation below the mean (about the 31st percentile) increases the student's standing by 15 percentile ranks.

For students scoring in the bottom quarter on the pre-test, the effect of the treatment on the total holistic scores is estimated to be half a standard deviation or more. Although the effects are largest at the lowest levels of the pre-test, statistically significant differences were found between total holistic scores for treatment and control group students with pre-test scaled scores up to 588 -- the 74th percentile in the total group.

Tables 3, 4, and 5 provide illustrations for the three separate holistic scores -- illustrations that parallel the one given for total scores in Table 2. Although there are minor differences, each of these separate scores exhibits the same pattern of findings that was exhibited by the total score, and each of these sets of findings supports the same general conclusion -- second-graders with relatively low initial scores on the language arts sections of a standardized test appear to profit a great deal from the Mason Thinking and Writing lessons in terms of the overall judged quality of their responses to brief writing tasks. The nearly identical pattern of results for reliable measures of judged overall quality of three separate writing exercises gives us a great deal of confidence in this conclusion.

Number of Words, Percent of Uncommon Words, and Number of Sentences

The Mason lessons do not encourage children to respond at length to writing exercises. In fact, some aspects of the program encourage succinct writing; for example, one of the colored cirles used by children in checking their own work reminds them to "stick to the point." The instructional program emphasizes use of complete and correct sentences, clarity, and order, rather than amount of written work produced. Thus, our analyses of numbers of words and sentences used by children in their responses to the three writing exercises are concerned with possible side effects of the program, rather than with effects the program is designed to produce.

Table 6 gives the results for number of words and Table 7 for number of sentences. The treatment/pre-test interaction was significant in both these cases and Johnson-Neyman analyses indicate that (within the effective range of pre-test scores) significant differences exist only for students with quite low pre-test scores -- from about the 20th percentile down. Moreover, the size of the differences (even in this region) appear to be small in absolute terms. In general, those students who experienced the Mason program did not appear to respond at greater length on the post-test exercises. The results suggest, however, that for students with very low pre-test scores, the regular practice in writing provided by the Mason lessons may result in somewhat longer responses to expository writing assignments.

In Table 8, results are shown for percentage of all words used in responses to the three exercises that were "uncommon" words. For our purposes, common words were defined as ones on the list of 200 words

given by Sundbye, Dyck, and Wyatt (1979) -- a list derived from the Kucera-Francis word list and the Harris-Jacobsen Short Readability Word List. We expected that, although vocabulary development is not a special focus of the Mason program, the regular practice in writing and encouragement provided by the Mason lessons would result in use of somewhat richer vocabulary in the writing of students in the treatment group. The results in Table 8 appear to bear out this theory, with statistically significant treatment/control differences for students scoring up to the 73rd percentile on the pretest. The largest difference shown in the table (a difference of 4.9 percent at the 10th percentile on the pretest) does not appear on its face to be a particularly large increase. However, since all of the most common articles, prepositions, and pronouns are on the list of common words used, a substantial percentage of words in most texts will be classified as common words, and the variation of percent of uncommon words is relatively restricted (e.g., it would never approach 100 percent). In fact, the difference of 4.9 percent was about three-quarters of the standard deviation for this variable observed in the control group, and it is a large difference in these terms.

Sentence Quality

We measured "sentence quality" by counting the number of correct sentences (complete sentence with initial capital and end puncuation) and dividing by the total number of sentences written to arrive at a percentage. Encouraging children to regularly use such complete sentences in their writing is a major aim of the Mason program. It is one of the features that children are to check in their own work when they

complete the writing exercises that make up part of the Mason lessons. Thus, we could expect to see large differences between the treatment and control groups if the lessons are effective in instilling skills and habits that would carry over to the post-test situation.

The results, which are given in Figure 2 and Table 9, show very large positive treatment effects, largest for students with low pre-test scores, but judged to be statistically significant for students from about the 90th percentile down. This outcome measure exhibits the largest effects (relative to control group variation) of any of the post-tests, and the results suggest that the Mason lessons have a strong effect on use of complete sentences in writing by second-graders in classes completing the program.

Spelling

Spelling was measured in several ways. One of the CTBS sections used as a post-test was the CTBS Spelling Test. We also counted the number of words correctly spelled in the student responses to the three writing exercises and calculated the percentage of all words spelled correctly. Similarly, we calculated percentages spelled correctly for common words (on the list of 200) and for uncommon words. The results of the analyses are given in Table 10 (CTBS Spelling), Table 11 (percent of all words spelled correctly), Table 12 (percent of common words spelled correctly), and Table 13 (percent of uncommon words spelled correctly).

The results of the analyses of these dependent variables are generally consistent. For students with low pre-test scores the treatment

has statistically significant positive effects on CTBS Spelling scores, on percent of all words spelled correctly, and on percent of common words spelled correctly. However, the effects are not very large, at least in comparison to those found for other outcomes. The estimated effects for students at the pre-test 10th percentile are 17 points for CTBS spelling (a difference amounting to 0.25 standard deviations) and 3.0% for percent of all words spelled correctly (a difference amounting to 0.30 standard deviations). Somewhat larger effects are found for percentage of common words spelled correctly, perhaps because the regular writing called for in the Mason program gives students considerable practice in spelling these frequently occurring words.

Summing Up -- Writing Scores

The following table gives the estimated treatment effects (statistically signficant effects only) in control group standard deviation units for students initially scoring at the 10th, 25th, and 50th percentiles for several writing measures:

	CTBS Pre-test Percentile		
	10th	25th	50th
Overall Judged Quality of Exercise Responses	.70	.50	.30
Use of Correct Sentences in Exercise Responses	.74	.60	.45
Use of Uncommon Words in Exercise Responses	.76	.56	.37
Spelling			
Common Words in Exercise Responses	.55	.42	.30
All Words in Exercise Responses	.30	.23	.16
CTBS Spelling Post-test	.25	.15	--
Number of Words in Exercise Responses	.35	--	--
Number of Sentences in Exercise Responses	.34	--	--

2. Results of Analyses of Student Data -- Content Indicators.

Since the content indicators took only values of 1 (present) or 0 (absent), logistic regression models were used to test for treatment effects while controlling for pre-test scores. Contrary to the regression results for writing scores described above, we did not find strong evidence for a treatment/pre-test interaction in these data, and thus our estimates of treatment effects are those obtained for models in which the logistic regressions of outcomes on pre-test scores are assumed to have the same slope for each group.

It should be noted, however, that these models fit a non-linear function of the success probabilities. Thus, even though no interaction term was included, estimated treatment/control differences in these probabilities based on the logistic models will not be constant across the range of pre-test scores. And, in fact, when the logistic regressions are graphed (see Figures 3 through 6), estimated control/treatment differences in these probabilities are smaller for students with high pre-test scores than for students with low pre-test scores on several variables.

We examined six indicators that resulted from content analyses of the student papers. Three of these indicators may be thought of as measures of "correctness" in completing the assigned task:

-- Were two or more interpretable similarities and two or more interpretable differences given for cats and dogs?

-- Were three or more steps given in order for washing your face?

-- Were clear rationales (common features) stated for each of two toy groups?

Preliminary analyses indicated that treatment group students were more likely than control group students to interpret the "how do you wash your face" exercise as one calling for a response listing a set of specific steps, probably because specific practice the students had received in the Mason lessons in giving written directions would have encouraged students to interpret the exercise in this way. A number of papers written by control group students gave acceptable answers to the exercise which did not involve a series of steps. Therefore, we limited our analyses of the content indicator for this exercise (three or more steps in order) to those students from each group who had in fact answered by listing one or more specific steps. As a result, 6 of 197 treatment group students and 81 of 186 control group students were excluded from analyses involving this variable.

For the three indicators of "correctness" listed above, we estimated the parameters of logistic regression models including the CTBS pre-test as a covariate and main effects for the treatment as well as a constant term.

We also examined three measures of specificity in responding to the exercise about cats and dogs:

-- Were superordinate categories (e.g., "they are both animals") rather than specific attributes used as the basis of a similarity?

-- Were one or more general differences cited?

-- Were specific attibutes used as the basis of one or more differences?

In the analyses of specificity indicators, we also included as independent variables in the logistic regression models total number of similarities provided (for the first of these indicators) or total number of differences provided (for the last two indicators), since any of these characteristics was more likely to be found in a paper with a large number of similarities or differences than in one with a small number. By including effects for the number of similarities or the number of differences in these models, we are able to examine whether specific or general similarities or differences are present more often in papers written by treatment or control group students, while controlling statistically for numbers of similarities or differences used. In the figures and tables illustrating group differences for these indicators, we have assumed papers stating either two similarities or two differences, as appropriate.

Results for the "correctness" indicators are given in Tables 14, 15, and 16 and in Figures 3, 4, and 5. For all three indicators, a positive treatment effect is evident, although the test statistic for the treatment effect for the first of the indicators (Table 14) is not quite significant at the 0.05 level for a two-tailed test (a value of 1.93 vs. a critical value of 1.96). Success in each case is also clearly related to initial standing on the pretest.

Results for the "specificity" indicators are given in Tables 17, 18, and 19, and the results for the last of these is graphed in Figure 6. Treatment effects are significant in all three cases and the treatment/control differences are in the expected directions -- indicating that in responding to the post-test exercises on similarities and differences of cats and dogs, students in classes completing the Mason lessons are less likely to use general rules as a basis for classification, and are more likely to cite specific attributes.

3. Responses to the Teacher Questionnaire.

At the end of the post-testing in December, participating teachers in both treatment and control classes were asked to complete a brief questionnaire. While several questions dealt with conditions that might affect test results and similar matters, responses to several questions concerning perceptions of the Mason program complement the information we have obtained from analyses of the student data.

As noted earlier, about two-thirds of the teachers of classes assigned to the treatment group, nearly all of whom had taught the Mason program in prior years as well, said that they believed that the lessons are more effective when taught during the course of a full year, rather than concentrated in a half-year as they were for this study.

Teachers in both groups, who had taught the program in this or in previous years, were asked for their judgment of the effects of the Mason program in relation to the regular curriculum. Specifically, they were asked whether the program had little effect, a moderate effect, or a strong effect on each of 17 skills or attitudes. Responses to this

question are summarized in Table 20. With one exception, the responses were highly positive, with from 72 percent to 100 percent of those responding stating that the program has a moderate or strong effect on the given skill or attitude in relation to the regular curriculum. The exception was spelling for which 59 percent of the respondents marked little effect.

These responses are generally consistent with the results of our analyses of the student data and our findings that students in the treatment group, particularly initially low scoring students, exhibit higher performance than control group students across a range of writing and thinking tasks. The relatively low ratings for spelling are also somewhat consistent with results of the analyses of student data, where we tended to find smaller treatment effects than were found for a number of other variables. In analyses of student data, however, we did find statistically significant differences favoring the treatment group for several spelling measures.

Finally, we asked teachers from both groups whether they felt that "the Mason program (in this year or in past years) has had an effect on you as a teacher or has changed the way you teach language arts or other subjects. Please comment." Fifteen of twenty teachers responding to the questionnaire wrote in positive comments -- many of which give additional insight into teacher perceptions and acceptance of the program. Portions of these comments are quoted below.

a. General comments about teaching of writing

"This program has helped me become more aware of different skills needed for writing."

"I put more emphasis on the writing part of language instead of the mechanics."

"I am more aware of the importance of expository writing."

" . . . It has caused me to stress more consistently the need for correct sentence structure and related items taught in the program."

"It has given me a handle in teaching such things as complete sentences, capitalization, punctuation. sticking to the point, etc. I have taught before, but the Mason program is the clearest and most comprehensive method I have used . . ."

b. Comments about use of explicit directions

"Yes. I feel I am more specific in my instructions and more thorough in teaching language."

" . . . I am also more cautious in giving explicit directions for any assigned task."

c. Comments about use of specific techniques in other areas

"Being able to remind children of certain skills by merely mentioning a colored circle has facilitated teaching in all areas."

"I transfer editing marks such as the 'x' for indenting and colored circles to other subjects."

" . . . I find myself constantly referring to the three ideas behind the circles when they apply to other written assignments . . ."

"Yes. I tend to emphasize skills taught in the Mason program in other subject areas."

d. Comments about enjoyment and humor

"It has increased my excitement and humor . . ."

"The Mason program re-emphasizes the value of play at the second grade level -- a fact many of us tend to forget as children grow (apparently) more sophisticated in modern society."

"I enjoy teaching the program and have fun with it. Because of that, perhaps it has had an effect on my teaching style."

"I enjoy teaching it -- it uses children's imaginations in helping others. It's a good reinforcer to what is taught in our curriculum."

REFERENCES

Fowles, M. E., Basic Skills Assessment: Manual for Scoring the Writing Sample, Princeton, New Jersey: Educational Testing Service, 1978.

Huitema, B. E., The Analysis of Covariance and Alternatives, New York: John Wiley & Sons, 1980.

Johnson, P. O., and Neyman, J., Tests of certain linear hypotheses and their application to some educational problems. Statistical Research Memoirs, 1936, 1, 57-93.

Mason Early Education Foundation, Thinking and Writing: Building a Foundation for Expository Writing, Second Grade Level, Princeton, New Jersey: Author, 1981.

Strickland, D. S., Language for Daily Use, New York: Harcourt, Brace, Jovanovich, 1983.

Sundbye, N. W., Lyck, N. J., and Wyatt, F. R., Essential Sight Words Program, Allen, Texas: DLM Teaching Resources, 1979.

APPENDIX

The following description of the aims and features of the Thinking and Writing program, second grade level, was provided by the program developer.

FOCUS

The focus of the second grade expository writing program is on the fundamental skills of composing and writing concise, effective sentences.

THINKING SKILLS

The thinking skills that the program aims to develop in the second grade are:

Classification

- -- of objects according to their shared characteristics
- -- through the use of specific and general words
- -- through generalizing about specific things
- -- according to fact and opinion
- -- according to fact and make-believe
- -- in which the view of the classifier determines whether a thing may or may not fit into one category or more than one category

Sequence

-- in time which is absolute

-- in time which is not absolute

-- in stories (sequence of events)

-- in directions or instructions

Relevance

-- distinguishing between the necessary and the unnecessary

-- distinguishing between the relevant and the irrelevant

WRITING SKILLS

Expository writing skills for the second grade level are developed in day to day lessons. These lessons give the students practice - first in thinking about what they want to say, then in putting their thoughts into words, and finally in writing them in a concise sentence.

WRITING MECHANICS

The second grade program also provides an introduction to the mechanics of writing:

-- the composition of a sentence - it begins with a capital letter, generally ends with a period, and is a complete sentence.

-- the composition of a paragraph - it consists of two or more related sentences. The first sentence is indented. The other sentences follow one after the other.

UNUSUAL FEATURES OF THE PROGRAM

Brevity in Writing

We recognize that at the second grade level the drawing of each letter is often a laborious procedure. For this and other reasons we limit the writing in the expository writing program to one or a very few sentences.

Another reason for requiring brevity, quite apart from the purely mechanical aspect, is to encourage the students to develop their ability to think, to organize and condense their thoughts, and then to express them - not in a page or pages of undisciplined writing, but in a few effective sentences, sometimes in only one sentence.

A third reason for having the students write only a few sentences is to allow them to both write and correct the work in a short period of time while interest in completing the project with care and attention to detail can be sustained.

Unique Incentives

A number of unique incentives are used in teaching this program at the second grade level -- among them, a storyline and a color code system.

The amusing imaginary "Mr. Magargle" provides the class with an incentive for clear thinking and writing. Though always well-meaning, Mr. Magargle is absent-minded and constantly requires the children's help. The children work together to meet his needs, not to compare

their individual writing abilities or to please the teacher. This helps to give the pupils affirmative attitudes toward themselves as effective writers and it helps them to establish a good self image.

The storyline also provides the incentive for the children to practice focusing their attention willingly and even eagerly on listening and following directions.

The color code system provides the children with an incentive to correct their own work. The color rings remind the students of what needs to be checked:

-- is each sentence a complete sentence?

-- does the composition stick to the point?

-- are the sentences written in the correct sequence?

Introduction of the Summary Sentence

The students are introduced to the composition of a summary sentence in tiny steps, week after week throughout the school year. In order to foster and maintain the pupils' interest in this work, the teacher begins by composing the summary sentence herself. Then she allows individual students to volunteer their suggestions. Finally the class as a group is allowed to compose the summary sentence.

Self Correcting Feature

The self-correcting feature, which is a part of each expository writing lesson, teaches the pupils independence in their writing. It requires them each, individually, to review and correct their own work -- to give their attention to the mechanics and spelling involved and to

make sure that their writing puts clearly into a sentence or two what they truly want to say.

THE PURSUIT OF EXCELLENCE

We have focused our lessons on a number of practical measures that help students to pursue excellence and to want to pursue it.

TABLES

Table 1

DESCRIPTIVE STATISTICS

	Mean	Standard Deviation	Minimum	Maximum
Pre-test Scaled Scores				
Avg. CTBS Language	541.83	67.44	364.00	675.67
Post-test Scaled Scores				
CTBS Spelling	583.23	61.04	360.00	642.00
CTBS Language Mechanics	595.21	43.58	443.00	695.00
CTBS Language Expression	603.37	87.78	278.00	715.00
Writing Quality Scores				
Holistic -- Cats & Dogs	5.76	1.44	2.00	8.00
Holistic -- Why We Write	4.51	1.67	2.00	8.00
Holistic -- Wash Your Face	5.15	1.56	2.00	8.00
Holistic -- Total	15.42	3.81	6.00	24.00
Exercise Counts				
Number of Words	61.37	22.52	10.00	154.00
Percent Uncommon Words	36.72	6.76	20.00	73.53
Pct Spelled Ok -- All	88.28	9.21	39.29	100.00
Pct Spelled Ok -- Common	93.51	7.12	52.38	100.00
Pct Spelled OK -- Uncommon	79.19	17.41	0.00	100.00
Number of Sentences	8.06	2.59	1.00	20.00
Pct Correct Sentences	73.50	25.04	0.00	100.00
Content Indicators				
Two Sim & Diff -- Cats/Dogs	.57	.50	0.00	1.00
Superord. Sim -- Cats/Dogs	.21	.41	0.00	1.00
General Diff -- Cats/Dogs	.34	.48	0.00	1.00
Specific Diff -- Cats/Dogs	.71	.46	0.00	1.00
3+ Steps in Order -- Wash Face	.87	.34	0.00	1.00
Both Reasons OK -- Toygroups	.76	.43	0.00	1.00

The number of cases for all variables is 383 with the exception of the indicator of 3+ Steps in Order -- Wash Face, for which the number of cases is 296. For cases with 0 steps (81 members of the control sample and 6 members of the treatment sample), this variable was considered to be undefined.

Table 2

ESTIMATED EFFECTS FOR TOTAL HOLISTIC SCORES

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	10.65	13.55	2.90 **	.70
25	12.91	14.97	2.06 **	.50
50	15.13	16.37	1.24 **	.30
75	16.68	17.35	0.67	--
90	18.14	18.27	0.13	--

Maximum Pre-test Score Associated
with Statistically Significant Effect
($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 588

Total Group
Percentile: 74

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0397	0.0250	-3.27 **
Constant	-6.7239	2.5998	

Multiple R = .643

---

* $P < 0.05$
** $P < 0.01$

Table 3

ESTIMATED EFFECTS FOR HOLISTIC SCORES -- CATS & DOGS EXERCISE

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	4.09	5.10	1.01 **	.62
25	4.88	5.60	0.72 **	.44
50	5.66	6.09	0.44 **	.27
75	6.20	6.42	0.23	--
90	6.71	6.76	0.05	--

Maximum Pre-test Score Associated
with Statistically Significant Effect
($P < 0.05$, Johnson-Neyman Analysis)

Scaled Score: 583

Total Group
Percentile: 71

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0139	0.0088	-2.86 **
Constant	-1.9979	1.2541	

Multiple R = .594

---

* $P < 0.05$
** $P < 0.01$

Table 4

ESTIMATED EFFECTS FOR HOLISTIC SCORES -- WHY WE WRITE EXERCISE

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	3.08	3.97	.89 **	.51
25	3.79	4.37	.58 **	.33
50	4.48	4.76	.28	--
75	4.97	5.03	.06	--
90	5.43	5.29	-.14	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 546

Total Group Percentile: 48

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0.25	0.0070	-2.36 *
Constant	-2.3765	0.8958	

Multiple R = .438

---

* $P < 0.05$
** $P < 0.01$

Table 5

ESTIMATED EFFECTS FOR HOLISTIC SCORES -- WASH YOUR FACE EXERCISE

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	3.48	4.49	1.01 **	.63
25	4.24	5.01	0.77 **	.47
50	4.99	5.53	0.54 **	.34
75	5.51	5.89	0.38 *	.24
90	6.00	6.23	0.23	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 599

Total Group Percentile: 79

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0133	0.0092	-2.05 *
Constant	-2.3494	0.4499	

Multiple R = .553

---

* $P < 0.05$
** $P < 0.01$

Table 6

ESTIMATED EFFECTS FOR NUMBER OF WORDS

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	48	57	9 *	.35
25	55	60	5	--
50	62	63	1	--
75	67	65	-2	--
90	72	67	-5	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P < 0.05$, Johnson-Neyman Analysis)

Scaled Score: 486

Total Group Percentile: 22

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.1256	0.0527	-2.21 *
Constant	-7.2366	33.8844	

Multiple R = .302

---

* $P < 0.05$
** $P < 0.01$

Table 7

ESTIMATED EFFECTS FOR NUMBER OF SENTENCES

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	6.5	7.5	1.0 **	.34
25	7.3	7.9	0.6	--
50	8.2	8.2	0.0	--
75	8.8	8.4	-0.4	--
90	9.3	8.7	-0.6	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 484

Total Group Percentile: 21

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.1523	0.0060	-2.45 *
Constant	-0.2225	4.9118	

Multiple R = .313

---

* $P < 0.05$
** $P < 0.01$

Table 8

ESTIMATED EFFECTS FOR PERCENT UNCOMMON WORDS

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	33.8	38.7	4.9 **	.76
25	34.7	38.3	3.6 **	.56
50	35.6	38.0	2.4 **	.37
75	36.3	37.8	1.5	--
90	36.9	37.6	0.7	--

Maximum Pre-test Score Associated
with Statistically Significant Effect
($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 586

Total Group
Percentile: 73

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0163	-0.0058	-2.18 *
Constant	26.6747	41.2082	

Multiple R = .232

---

* $P < 0.05$
** $P < 0.01$

Table 9

ESTIMATED EFFECT FOR PERCENT CORRECT SENTENCES

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	47.9	68.9	21.0 **	.74
25	58.3	75.2	16.9 **	.60
50	68.6	81.3	12.7 **	.45
75	75.7	85.6	9.9 **	.35
90	82.5	89.6	7.1 *	.25

Maximum Pre-test Score Associated with Statistically Significant Effect ($P < 0.05$, Johnson-Neyman Analysis)

Scaled Score: 628

Total Group Percentile: 90

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.1833	0.1096	-2.24 *
Constant	-32.4397	20.8937	

Multiple R = .517

---

* $P < 0.05$
** $P < 0.01$

Table 10

ESTIMATED EFFECTS FOR CTBS SPELLING POST-TEST

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	498	515	17 **	.25
25	544	554	10 *	.15
50	589	591	2	--
75	621	618	-3	--
90	650	643	-7	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P < 0.05$, Johnson-Neyman Analysis)

Scaled Score: 509

Total Group Percentile: 31

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.8046	0.6774	-2.44 *
Constant	145.9326	218.3645	

Multiple R = .831

---

* $P < 0.05$
** $P < 0.01$

Table 11

ESTIMATED EFFECTS FOR PERCENT OF ALL WORDS SPELLED CORRECTLY

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	77.8	80.8	3.0 **	.30
25	83.1	85.4	2.3 **	.23
50	88.3	89.9	1.6 *	.16
75	91.9	93.1	1.2	--
90	95.3	96.1	0.8	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P < 0.05$, Johnson-Neyman Analysis)

Scaled Score: 566

Total Group Percentile: 59

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0927	0.0809	-1.11
Constant	37.1898	45.3714	

Multiple R = .658

---

* $P < 0.05$
** $P < 0.01$

Table 12

ESTIMATED EFFECTS FOR PERCENT OF COMMON WORDS SPELLED CORRECTLY

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	86.3	90.8	4.5 **	.55
25	89.6	93.0	3.4 **	.42
50	92.8	95.2	2.4 **	.30
75	95.1	96.7	1.6 *	.20
90	97.2	98.0	0.8	--

Maximum Pre-test Score Associated with Statistically Significant Effect ($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: 593

Total Group Percentile: 76

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.0575	0.0384	-2.04 *
Constant	61.1540	74.0235	

Multiple R = .520

---

* $P < 0.05$
** $P < 0.01$

Table 13

ESTIMATED EFFECTS FOR PERCENT OF UNCOMMON WORDS SPELLED CORRECTLY

%ile Standing on Pre-test	Expected Post-test Values Control	Treatment	Difference	Difference/ S.D. (Control)
10	61.1	64.6	3.5	--
25	70.6	73.1	2.5	--
50	79.9	81.4	1.5	--
75	86.4	87.2	0.8	--
90	92.6	92.7	0.1	--

Maximum Pre-test Score Associated
with Statistically Significant Effect
($P \leq 0.05$, Johnson-Neyman Analysis)

Scaled Score: ---

Total Group
Percentile: --

Estimated Regression Coefficients (outcome on pre-test):

	Control	Treatment	T-value for Slope Difference
Slope	0.1670	0.1487	-0.87
Constant	-12.0884	-0.4821	

Multiple R = .622

---

* $P < 0.05$
** $P < 0.01$

Table 14

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- CORRECTNESS

Two Similarities and Two Differences -- Cats & Dogs

%ile Standing on Pre-test	Estimated Probability of Attribute Control	Treatment	Difference
10	.28	.37	.09
25	.41	.52	.11
50	.56	.66	.10
75	.66	.75	.09
90	.74	.81	.07

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	0.0106	6.01 **
Treatment	0.4254	1.93
Constant	-5.5954	

* P < 0.05
** P < 0.01

Table 15

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- CORRECTNESS

Three or More Steps in Order -- Wash Your Face

%ile Standing on Pre-test	Estimated Probability of Attribute Control	Treatment	Difference
10	.62	.84	.22
25	.72	.89	.17
50	.79	.93	.14
75	.84	.94	.10
90	.87	.96	.09

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	0.0076	2.88 **
Treatment	1.1840	3.33 **
Constant	-2.8394	

---

* P < 0.05
** P < 0.01

Table 16

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- CORRECTNESS

Common Features Given for Both Toy Groups

%ile Standing on Pre-test	Estimated Probability of Attribute		Difference
	Control	Treatment	
10	.54	.71	.17
25	.63	.78	.15
50	.71	.84	.13
75	.76	.87	.11
90	.80	.89	.09

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	0.0066	3.62 **
Treatment	0.7342	2.90 **
Constant	-2.7324	

---

* $P < 0.05$
** $P < 0.01$

Table 17

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- SPECIFICITY

Superordinate Similarity Used -- Cats & Dogs

Illustration assumes paper with two similarities:

%ile Standing on Pre-test	Estimated Probability of Attribute Control	Treatment	Difference
10	.19	.11	-.08
25	.23	.13	-.10
50	.27	.16	-.11
75	.31	.19	-.12
90	.34	.21	-.13

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	0.0042	2.09 *
No. of Similarities	0.4746	2.42 *
Treatment	-0.6457	-2.47 *
Constant	-4.2669	

---

* $P < 0.05$
** $P < 0.01$

Table 18

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- SPECIFICITY

General Difference Used -- Cats & Dogs

Illustration assumes paper with two differences:

%ile Standing on Pre-test	Estimated Probability of Attribute Control	Treatment	Difference
10	.52	.39	-.13
25	.50	.37	-.13
50	.48	.35	-.13
75	.47	.34	-.13
90	.45	.33	-.12

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	-0.0014	-0.79
No. of Differences	1.0268	5.02 **
Treatment	-0.5315	-2.33 *
Constant	-1.3558	

* P < 0.05
** P < 0.01

Table 19

ESTIMATED EFFECTS FOR CONTENT INDICATORS -- SPECIFICITY

Specific Difference Used -- Cats & Dogs

Illustration assumes paper with two differences:

%ile Standing on Pre-test	Estimated Probability of Attribute Control	Treatment	Difference
10	.60	.76	.16
25	.63	.79	.16
50	.66	.81	.15
75	.68	.82	.14
90	.70	.83	.13

Estimated Coefficients -- Logistic Regression Model

Variable	Coefficient	Coefficient/ Std. Error
Pre-test	0.0023	1.28
No. of Differences	0.3056	1.59
Treatment	0.7686	3.27 **
Constant	-1.2196	

---

* $P < 0.05$
** $P < 0.01$

Table 20

TEACHER JUDGMENTS OF EFFECTS OF MASON PROGRAM
IN COMPARISON TO THE REGULAR CURRICULUM

Skills/attitudes ranked according to percentage marking "strong effect."	Percent		
	Little Effect	Moderate Effect	Strong Effect
Getting ideas in order	0	17	83
Writing complete sentences	0	23	77
Following directions	0	23	77
Classification	6	19	75
Distinguishing between absolute and changeable order	0	33	67
Distinguishing between the necessary and the unnecessary	0	39	61
Distinguishing between specific words and general words	0	39	61
Positive attitude toward writing	6	33	61
Distinguishing between fact and opinion	0	44	56
Capitalization and punctuation	0	50	50
Distinguishing between fact and make-believe	22	33	44
Checking and self-correction	28	28	44
Fluency/ease in writing	0	61	39
Summarizing	6	56	39
Ability to write a clear, well-focused, very short paragraph on an assigned topic	6	59	35
Listening skills	6	61	33
Spelling	59	35	6

FIGURES

FIGURE 1

TOTAL HOLISTIC SCORE

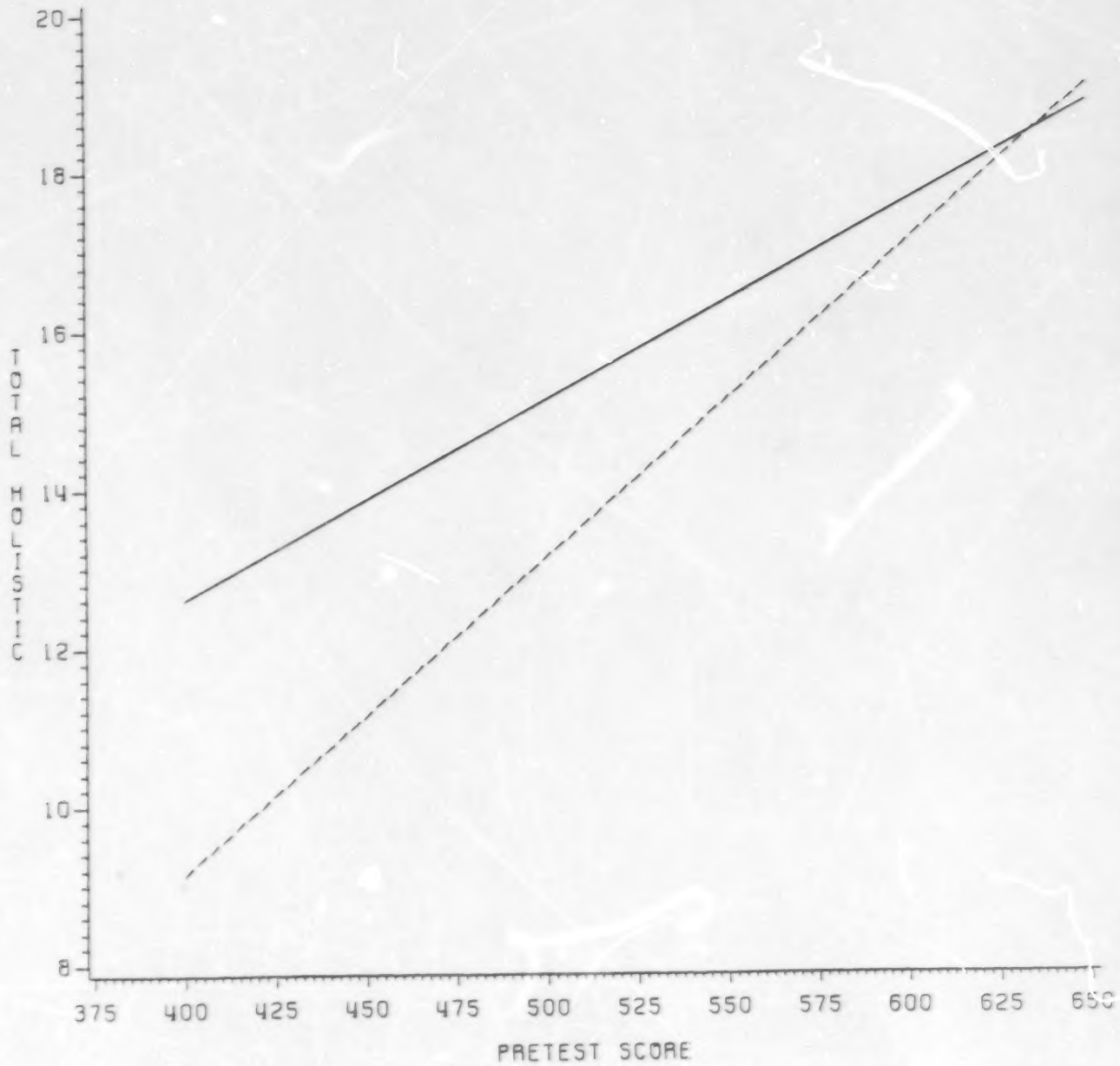


FIGURE 2

PERCENT CORRECT SENTENCES

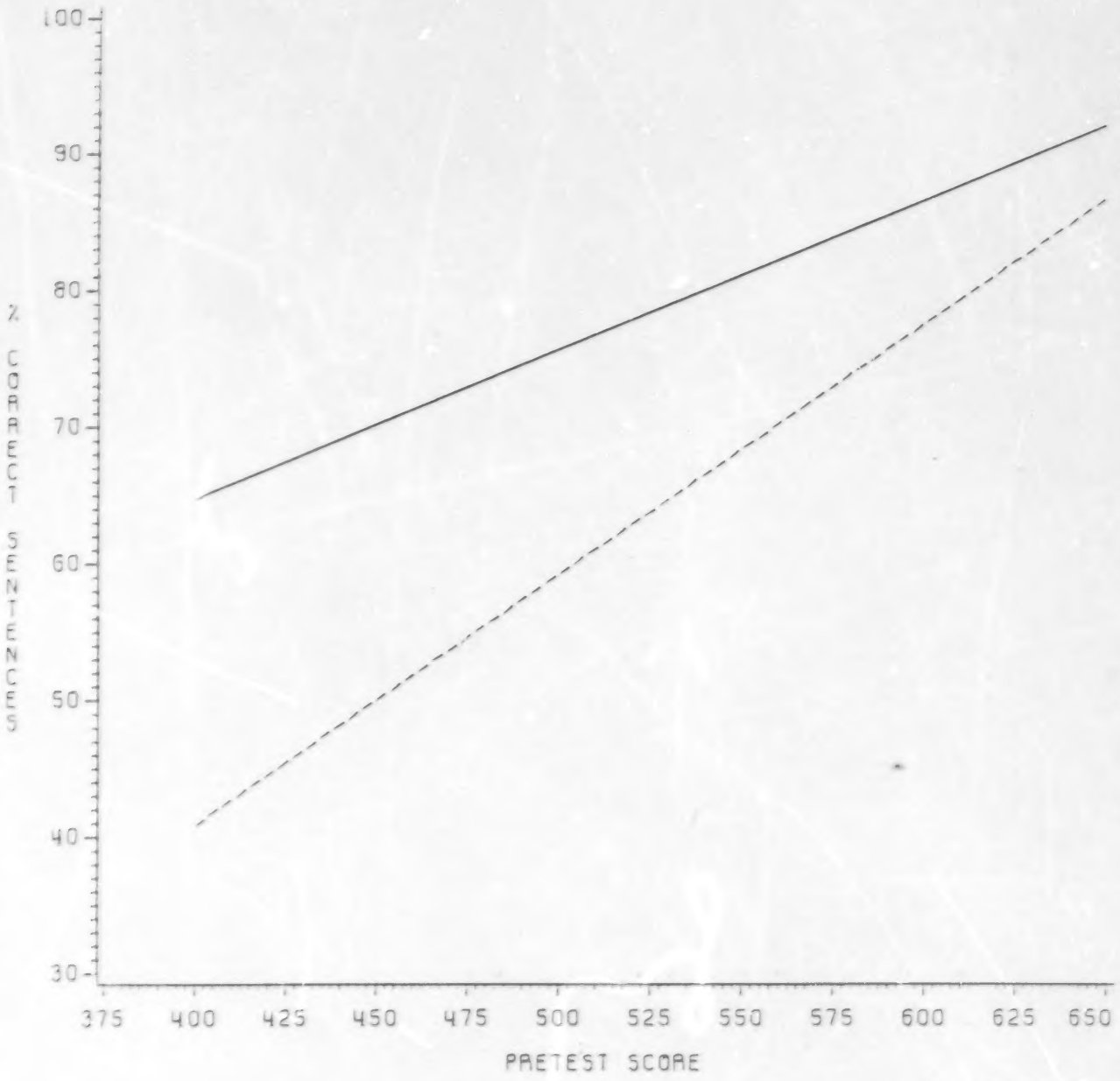
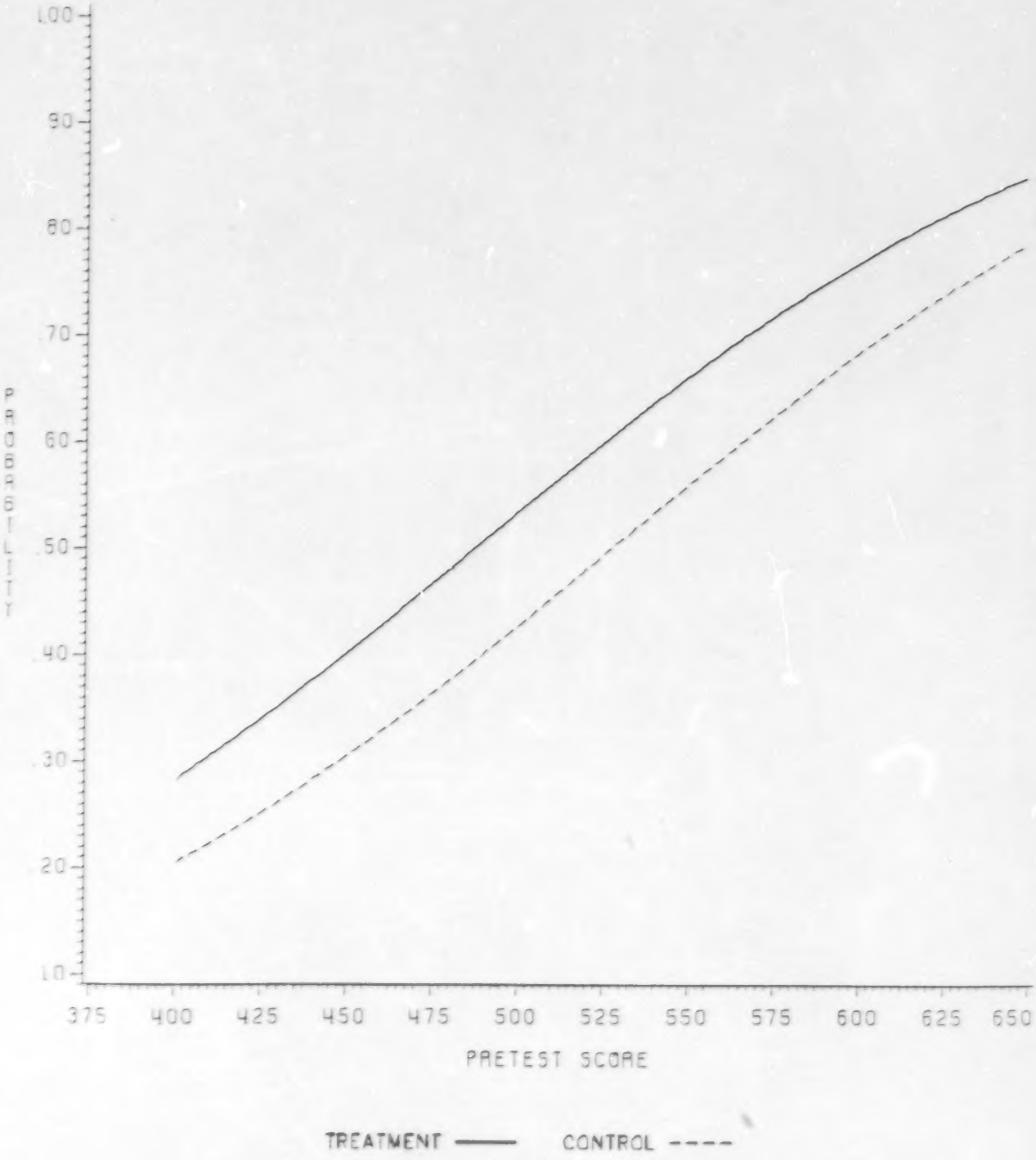


FIGURE 3

TWO SIMILARITIES AND DIFFERENCES -- CATS & DOGS



77

FIGURE 4

THREE+ STEPS IN ORDER -- WASH FACE

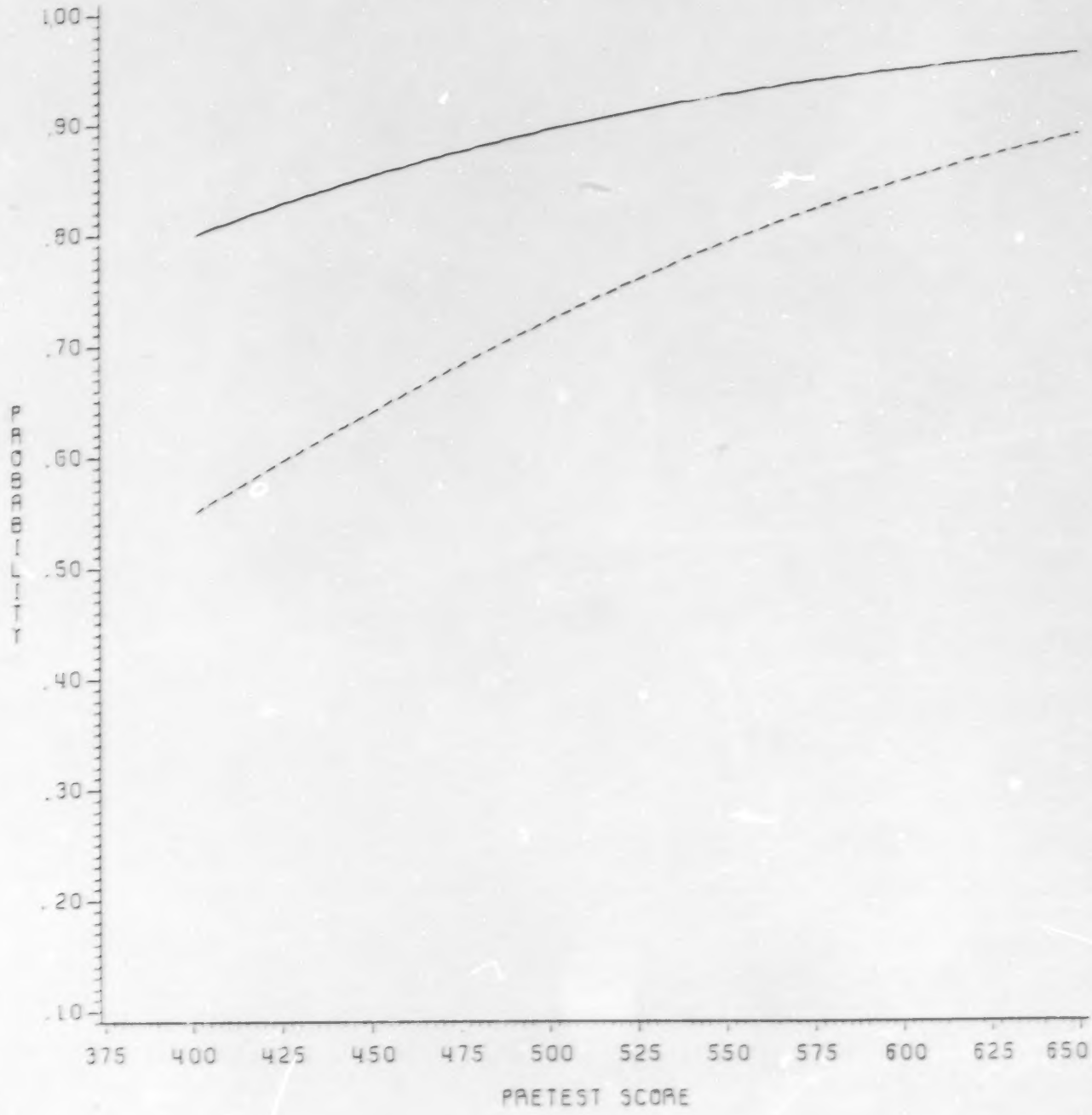


FIGURE 5

TWO GOOD RATIONALES -- TOYGROUPS

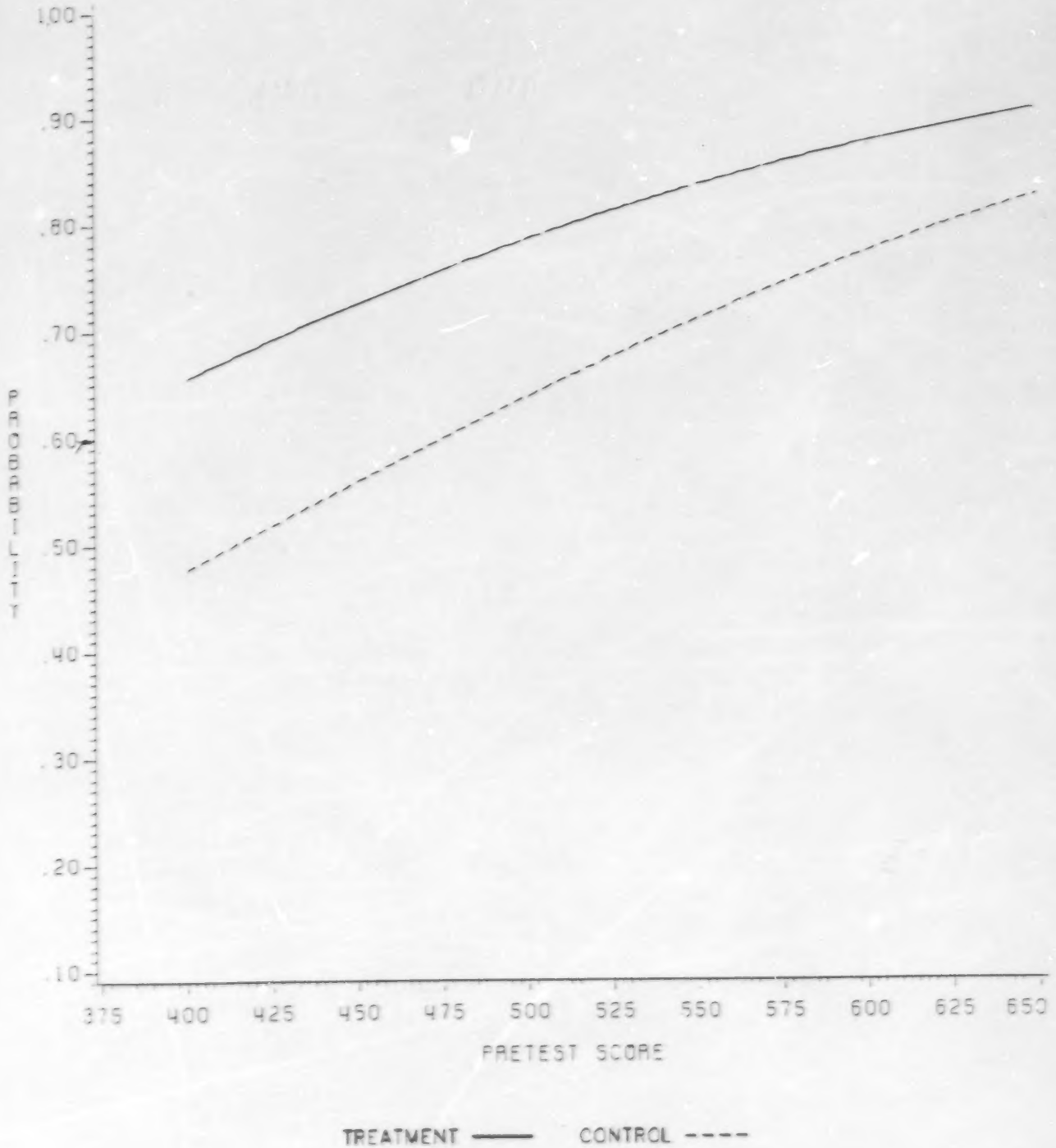


FIGURE 6

SPECIFIC DIFFERENCE USED -- CATS & DOGS